

An Anomalous TY1 Structure Attributed to an Error in Reverse Transcription

BEVERLY ERREDE,^{1*} MAHSHID COMPANY,¹ AND RONALD SWANSTROM²

Department of Chemistry¹ and Department of Biochemistry,² University of North Carolina, Chapel Hill, North Carolina 27514

Received 19 September 1985/Accepted 15 January 1986

We have determined the nucleotide sequence of both delta elements of a Ty1 transposon inserted near the *CYC7* gene in the *Saccharomyces cerevisiae* *CYC7-H2* mutant. The upstream delta element in this Ty1 has an unusual inverted repeat structure that may have been formed by an error during reverse transcription.

Ty transposable elements are a dispersed repeated gene family in *Saccharomyces cerevisiae* (3). Ty elements are similar to proviral forms of vertebrate retroviruses in structure (25), transcriptional properties (8), and gene organization (4, 18). Both proviruses and Ty elements are bounded by terminal repeats, called long terminal repeats or LTRs in retroviruses and delta elements in Ty1. The terminal repeats flank open reading frames which, in proviruses, encode gene products involved in viral replication and, in Ty elements, encode gene products presumed to be involved in transposition. Transcription from both proviruses and Ty elements initiates in the upstream copy of the terminal repeat and terminates in the downstream copy. Because of an overlap in the positioning of the initiation and termination sites within the terminal repeat sequence, the transcripts from both Ty1 and proviruses have a short terminal repeat. Finally, sequences analogous to those used in the priming of viral DNA synthesis are also present in Ty elements. These are a tRNA binding site at one internal boundary of the terminal repeat (7, 24) and a polypurine tract at the other internal boundary (25). A comparison of the structures of Ty1 elements, proviruses, and their transcripts is shown in Fig. 1.

It has recently been shown that transposition of Ty1 occurs by a mechanism involving reverse transcription of Ty1 RNA (2). In addition, evidence demonstrating a Ty-encoded reverse transcriptase and Ty-related viruslike particles has been presented (13, 19). These observations draw additional parallels with retroviruses, which generate the DNA form of their genome by reverse transcription of genomic RNA. Because of the similarities in the structures of both the DNA and RNA forms of these elements, it is likely that the steps involved in transposition by Ty1 are very similar to the well-characterized steps involved in the synthesis of retroviral DNA. In characterizing the *CYC7-H2* Ty1 insertion mutation, we observed an unusual structure for a delta element (10). Further characterization of this element has led us to conclude that the unusual structure may have been generated by a mistake during reverse transcription.

The *CYC7-H2* mutation in *S. cerevisiae* resulted from insertion of a Ty1 element into the noncoding region of the *CYC7* structural gene (9). We previously determined 1.5 kilobases (kb) of DNA sequence from the 5' end of the *CYC7-H2* Ty1 (10). One unexpected result was the finding that the 5' delta element is only 136 base pairs (bp) long compared to the normal size of approximately 330 bp (Fig.

2A). The truncated element has no sequences that correspond to the U3 domain of a conventional delta sequence as defined in Fig. 1. In an attempt to understand the origin of the defect in the 5' delta element, we determined the nucleotide sequence of the 3' delta element of the *CYC7-H2* Ty1 and compared it with the sequence of the truncated 5' element.

Restriction fragments a-b, a-c, and b-c, which include sequences from the 3' delta element of the *CYC7-H2* Ty1, were cloned into M13 vectors (Fig. 3). The nucleotide sequence of the resulting clones was determined by the chain termination method (1, 21, 22). The sequence of the 3' delta element is shown in Fig. 2A compared with the sequence of the truncated 5' delta element from the same Ty1 and with the 5' delta element from Ty912 (4, 11) as representative of a typical delta sequence. Several points can be made based on this sequence comparison: (i) the 3' delta from *CYC7-H2*

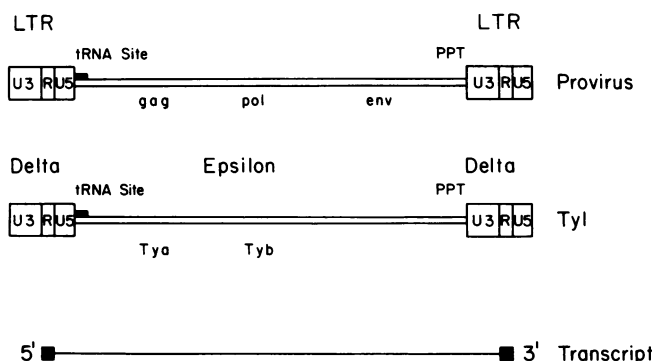


FIG. 1. Schematic structures for a typical provirus, a Ty1 element, and their transcripts. The important structural and genetic domains are indicated by the following notations: LTR, long terminal repeat sequence of proviruses; Delta, long terminal repeat sequence of Ty1 elements; Epsilon, internal domain of Ty1 elements; U3, sequence at the 3' terminus of the RNA transcript that is present in the long terminal repeat (excluding R); U5, sequence corresponding to the terminal redundancy at each end of the RNA transcript (solid box); U3, sequence at the 3' terminus of the RNA transcript that is present in the long terminal repeat (excluding R); U5, sequence corresponding to the terminal redundancy at each end of the RNA transcript (solid box); tRNA site (solid bar), binding site for the tRNA primer used to initiate viral DNA synthesis; PPT, polypurine tract, the probable primer site for synthesis of the second strand of viral DNA; gag, coding region for capsid proteins; pol, coding region for reverse transcriptase and endonuclease; env, coding region for envelope glycoproteins; Tya, first open reading frame protein of Ty1; Tyb, second open reading frame protein of Ty1.

* Corresponding author.

A

20
CYC7 atgttgtgtgcctttgtcctt..... CYC7-H2 5' DELTA
PPT
EPSILON ctattacattatgggtggtaTGTGGAATAGAAATCAACTATCATCTACTAACTAGTATTACATTACTAGTATATTATC CYC7-H2 3' DELTA
|||||
TGTGGAATAGAAATCAACTATCATCTACTAACTAGTATTACATTACTAGTATATTATC TY912 5' DELTA

70 120
..... CYC7-H2 5' DELTA
ATATACGGTGTTAGAAGATGACGCAAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGGA CYC7-H2 3' DELTA
|||||
ATATACGGTGTTAGAAGATGACGCAAAATGATGAGAAATAGTCATCTAAATTAGTGGAAGCTGAAACGCAAGGATTGATAATGTAATAGGA TY912 5' DELTA

170 220
.....
TGAGAAATGGGTGAATGTTGAGATAAATTGTTGGGATTCC CYC7-H2 5' DELTA
|||||
TCAATGAATATAAACATATAAAACGGAATGAGGAATAATCGTAATATTAGTATGTAGAAATATAGATTCCATTTTGAGGATTCCCTATATC CYC7-H2 3' DELTA
|||||
TCAATGAATATAAACATATAAAATGATGATAATAATATTATAGAAATT...GTGTAGAATTGCAGATTCCCTTTTATGGATTCCCTAAATC TY912 5' DELTA

RNA Start 270 320
ATTGTTGATAAAGGCTATAATATTAGGTATACAGAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTAATTACCCACAAA CYC7-H2 5' DELTA
|||||
CTCGAGGAGAACTTCTAGTATATTCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTAATTACCCACAAA CYC7-H2 3' DELTA
|||||
CTTGAGGAGAACTTCTAGTATATTCTGTATACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCA TY912 5' DELTA
Poly A

1 RNA Site
TTTCTCAgggtagcgctgtgacttcgg EPSILON CYC7-H2 5' DELTA
|||||
TTTCTCAgtctcttggcccttggcgcca CYC7 CYC7-H2 3' DELTA
|||||
TTTCTCA TY912 5' DELTA

B

```
TACAGAAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTAATTACCCACAAATTTCTCA      CYC7-H2 5' DELTA      (270-337)
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
TACCTAATATTATAGCCTTTATCAACAATGGAATCCCAACAATTATCTCAACATTCACCCATTTCTCA      CYC7-H2 5' DELTA COMPLEMENT (269-202)
```

FIG. 2. Comparison of delta sequences. (A) The truncated 5' delta sequence from the *CYC7-H2* Ty1 is compared with the 3' delta sequence from the same Ty1 and with the previously published 5' delta sequence from Ty912 (4, 11). Delta sequences are shown in capital letters, and flanking sequences are indicated by lowercase letters. Dots are inserted in sequences as necessary to maintain alignment of corresponding regions. Tic marks indicate identical bases between two delta sequences on adjacent lines. Asterisks above the *CYC7* flanking sequence indicate the 5-base pair duplication of host DNA that occurred upon insertion of the Ty1 element (16). The polypurine tract (PPT) in epsilon DNA at the 3' delta boundary (25) is indicated. The presumed tRNA binding site in epsilon DNA at the 5' delta boundary, as identified by homology to the 3' end of the yeast initiator tRNA (7), is indicated. The major transcription start site (↓) for Ty RNA in the 5' delta element and the poly (A) addition site (↑) for Ty RNA in the 3' delta element (8) are indicated. The center of the inverted repeat is indicated by the long vertical line. The long horizontal arrows above the sequence of the *CYC7-H2* 5' delta element show the (imperfect) inverted repeat which extends from the center of the repeat to flanking DNA sequence. A smaller repeat in the normal delta element, centered at the same site, is shown by the short horizontal arrows above the other two elements. The smaller inverted repeat may have played a role in generating the larger inverted repeat. (See the text.) (B) The inverted repeat structure of the *CYC7-H2* 5' delta sequence compared by alignment of sequences reading away from the axis of symmetry. The *CYC7-H2* 5' delta sequence from position 270 to 337 in part A is shown on the top line. The complement of the *CYC7-H2* 5' delta sequence from position 269 to 202 in part A is shown on the bottom line. Tic marks indicate identical bases.

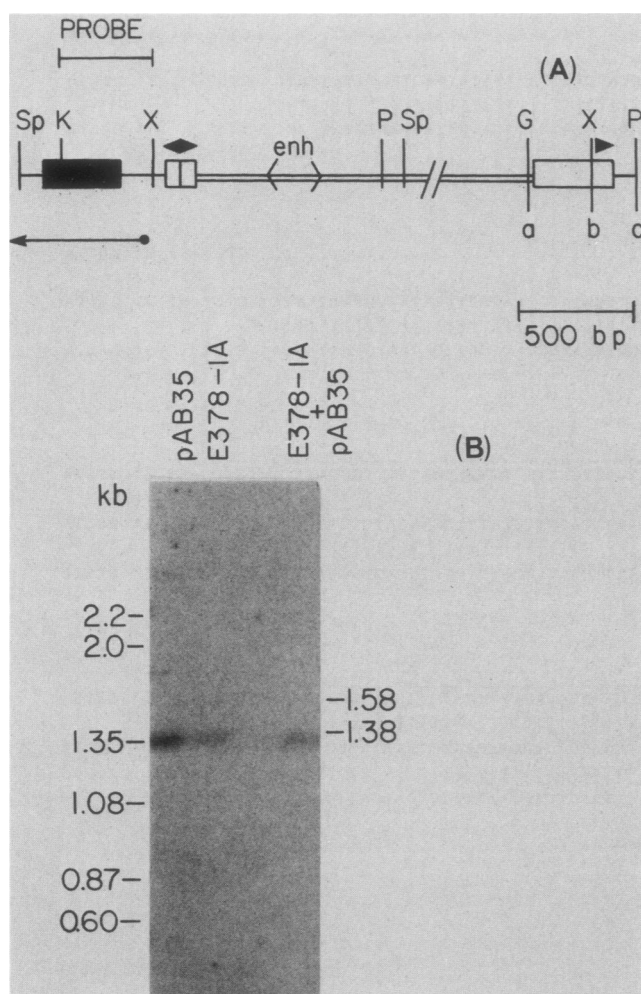


FIG. 3. Comparison of *CYC7*-Ty1 junction fragments from the cloned and genomic *CYC7-H2* gene. (A) Structure of the cloned *CYC7-H2* region. Solid box, the *CYC7* coding sequence; open box, the delta sequence; double line, the epsilon sequence; single line, the flanking sequence. The solid triangles above the diagram indicate the inverted repeat sequence in the 5' delta element (left) and the corresponding sequences in the 3' delta element (right). The region designated (enh) encompasses the enhancer region of the Ty1 element (10). The bar above the diagram indicates the 0.4-kb *KpnI-XhoI* fragment that was used as a hybridization probe for the Southern hybridization experiment shown in panel B. The arrow below the diagram indicates the 5'-to-3' direction and extent of *CYC7* transcription. Restriction sites shown are: *Bgl*II (G), *Kpn*I (K), *Pst*I (P), *Sph*I (Sp), and *Xho*I (X). (B) Autoradiogram comparing cloned and genomic *CYC7-H2* DNA. A 1- μ g sample of *Kpn*I- and *Pst*I-digested strain E378-1A genomic DNA was compared with 0.4 ng of *Kpn*I- and *Pst*I-digested pAB35 plasmid DNA. DNA samples were fractionated by electrophoresis on 1% agarose gels in TAE buffer (40 mM Tris, 5 mM sodium acetate, 1 mM EDTA, pH 7.9) at 1 V/cm. Conditions for Southern transfer and hybridization were according to procedures described in reference 17. Positions of size markers (lambda *Hind*III and ϕ X174 *Hae*III fragments) are indicated to the left of the autoradiogram. Marks on the right indicate positions expected for DNA fragments corresponding to either a 1.38-kb *Kpn*I-*Pst*I fragment encompassing a truncated delta element or a 1.58-kb fragment encompassing a conventional delta element.

Ty1 is normal; (ii) the insertion of the Ty1 element into the *CYC7* locus apparently occurred normally since both the normal and truncated delta elements are flanked by a 5-bp direct repeat of cellular DNA, as expected for a Ty1 transposition event (11, 12); (iii) the sequences in the truncated 5' delta element represent an inverted repeat centered at a position that is near the site that corresponds to the 5' end of Ty RNA (8); (iv) the center of the inverted repeat is precisely at the same point as a short, imperfect, inverted repeat that is present in normal delta sequences; and (v) the inverted sequence, which includes part of R and all of U5, is joined to cellular DNA. A direct comparison of sequences comprising the inverted repeat of the 5' delta element is made in Fig. 2B. Sequences extending from the center of symmetry to the epsilon junction (positions 270-337) are aligned with the complement of sequences extending from the center of symmetry to flanking cellular DNA (positions 269-202). The alignment illustrates the presence of a perfect inverted repeat for 43 bp beyond the short, imperfect repeat present in normal delta sequences. It also shows there are seven mismatches clustered near the end of the inverted repeat.

To show that the truncated delta element in the cloned DNA fragment from *CYC7-H2* was not generated during the cloning step, we compared the size of a restriction fragment that spans the truncated element both in the cloned DNA from the plasmid pAB35 (9) and in genomic DNA from the *CYC7-H2* strain E378-1A. Figure 3A shows a physical map of a portion of the *CYC7-H2* locus as determined from analysis of cloned DNA. When the cloned DNA was cleaved with *Kpn*I and *Pst*I, a 1.38-kb junction fragment between the *CYC7* coding region and the Ty1 element was released (Fig. 3B). If genomic DNA from the *CYC7-H2* locus contained a normal Ty1 element, then the junction fragment should have been 1.58 kb, a size difference that would have been resolved in this analysis. However, we found that the size of the junction fragment was the same in the genomic and cloned DNAs. This result indicates that the alteration in the 5' delta element occurred in *S. cerevisiae* before cloning.

The presence of a truncated delta element with an inverted repeat in an otherwise normal Ty1 element was an unexpected observation. An explanation for the origin of this structure is not immediately obvious. However, a plausible explanation can be found by extending the analogy between Ty1 elements and retroviruses.

The origin of the long terminal repeat structure in retrovirus DNA is intimately involved in the synthesis of viral DNA (14, 25). The steps involved in this process are outlined in Fig. 4A. Synthesis of the first strand of viral DNA is initiated on a genome-bound tRNA. Synthesis continues to the 5' end of the template through the U5 and R regions (Fig. 4A, step 1). The RNA template, in hybrid with DNA, is degraded by a DNA polymerase-associated RNase H activity, exposing the complement of the R sequence in the newly synthesized DNA. The complement of the R sequence in DNA pairs with the copy of the R sequence at the 3' end of the template, and synthesis continues along the length of viral RNA (step 2). The second strand of viral DNA is initiated at the U3 boundary of the first DNA strand (step 3). The regions U3, R, and U5 in the first strand are copied into duplex DNA, and synthesis continues into the first 15 to 20 bases of the tRNA primer, which is attached to the 5' end of the first strand. The first strand terminates near the 5' end of the RNA template after copying the tRNA binding site; the adjacent R and U5 regions in the template were previously degraded by RNase H when in hybrid with the initial DNA transcript. When the remaining RNA in hybrid is removed,

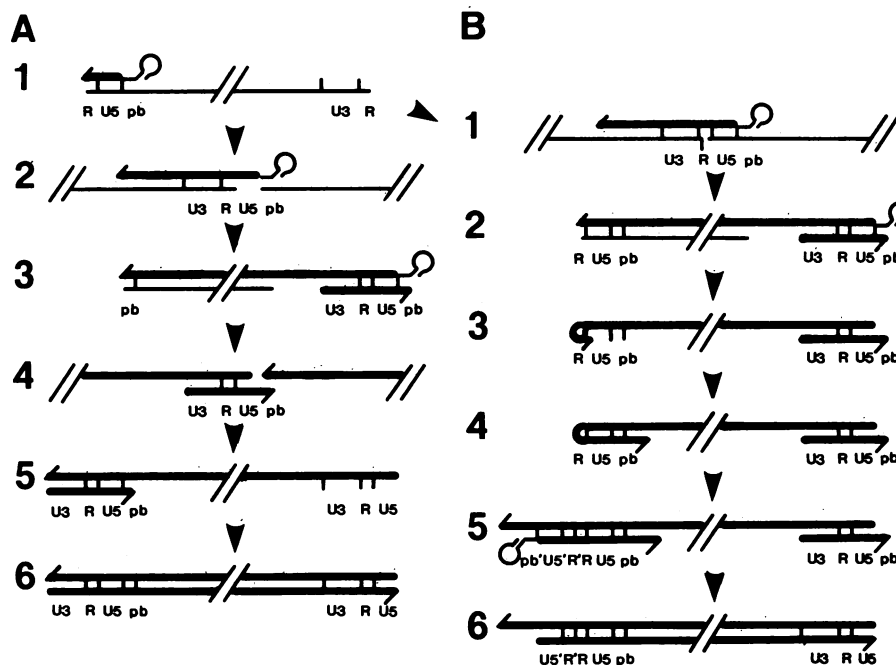


FIG. 4. Models for the synthesis of normal and aberrant terminal repeats. The thin line represents RNA, and the thick line represents DNA. The two ends involved in viral DNA synthesis can be contributed by circularizing a single viral subunit or by use of two viral subunits. The short looped structure represents the tRNA primer, with the primer binding site designated pb. U3, R, and U5 are as indicated in the legend to Fig. 1. R' and U5' are the complements of R and U5 when present on the same strand to indicate an inverted repeat. pb' designates the presence of a primer binding site sequence in the first strand of DNA. (A) Outline of the normal mode of retroviral DNA synthesis (14, 25). The individual steps are described in the text. (B) Proposed pathway for the synthesis of a viral long terminal repeat or yeast delta element with an inverted repeat of the R and U5 sequence. Each step is discussed in the text.

the 3' ends of the first and second strands have complementary tails corresponding to the sequences in the tRNA and tRNA binding site. These tails anneal (step 4), and synthesis continues with the first strand using the short second strand as template (step 5) such that the U3, R, and U5 domains are duplicated to form a direct repeat at the ends of viral DNA. The second strand is extended using the first strand as template to give a duplex linear molecule (step 6).

An alteration in the scheme used by retroviruses could account for the structure we observed in the Ty1 of the *CYC7-H2* mutant. Since the right-hand delta structure is intact, the error in synthesis may have occurred during the final steps of DNA synthesis. One possible scheme is shown in Fig. 4B. If RNase H digestion removed only a portion of the RNA from the 5' end of the RNA template after the initiation event, extension of a first strand could still occur (step 1), but synthesis would ultimately continue beyond the tRNA binding site and into a portion of the R region (step 2). This situation generates an intermediate with a normal right end but an aberrant left end in which the first strand is extended into R. The aberrant intermediate could not complete synthesis normally due to the lack of complementary tails involving tRNA sequences. Synthesis could continue if the 3' end of the first strand folded back on itself to form a hairpin (step 3). Synthesis could then continue into and beyond the tRNA binding site (step 4). This scheme accounts for the presence of an inverted repeat sequence in the initial single-stranded products. The process is similar to one described previously in which hairpin DNA synthesis at the 5' end of retroviral RNA was observed during DNA synthesis in vitro (23). There is a short, imperfect, inverted repeat (7 of 10 bases match) in the normal delta sequence which is centered at precisely the same point as the larger inverted

repeat in the truncated delta sequence. The presence of the short inverted repeat is consistent with the view that a hairpin could have formed in this region, permitting foldback synthesis (Fig. 2). According to this model, a perfect inverted repeat should be generated beyond the foldback region. Figure 2B shows there is a perfect inverted repeat for 43 bp followed by a clustering of seven mismatches near the end of the inverted repeat. Because of the clustering of mismatches it seems unlikely that these differences are the result of errors during reverse transcription. However, these differences could have arisen by repair or recombination events subsequent to the generation of the *CYC7-H2* Ty1 structure.

While the above scheme may account for the origin of the unusual delta structure, it does not account for the insertion of the aberrant element into host DNA. As judged by the direct repeat of cellular DNA flanking the element (Fig. 2), the insertion event was apparently normal. We can suggest two possible schemes for inserting this structure into the chromosome. First, after synthesis past the tRNA binding site, a second tRNA could be used, now bound to the DNA and priming second-strand synthesis from the left end of the molecule (Fig. 4B, step 5). This would provide a delta boundary sequence in duplex at each end of the molecule, although the left end would probably have a single-stranded tail (step 6). The second possibility is that the hairpin structure is used directly, perhaps by cleaving one strand (the foldback strand) at the duplex region of the delta/tRNA binding site boundary in a manner analogous to the cleavage of retroviral DNA by the viral endonuclease (5, 6, 15). The hairpin could then be unfolded to generate an end for insertion. Both of these schemes for insertion invoke a linear intermediate in insertion. However, a circular DNA could be

used in insertion, as with retroviruses (20), in which case the linear DNA structure would have to be repaired to a circle before insertion.

We thank John Olsen for assistance in preparing Fig. 4. We also thank Bradley Kosiba for discussing with us preliminary *CYC7-H2* sequence data.

This research was supported by Public Health Service research grants GM30619 (B.E.) and CA33147 (R.S.).

LITERATURE CITED

- Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**:3963-3965.
- Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink. 1985. Ty elements transpose through an RNA intermediate. *Cell* **40**:491-500.
- Cameron, J. R., E. Y. Loh, and R. W. Davis. 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell* **16**:739-751.
- Clare, J., and P. Farabaugh. 1985. Nucleotide sequence of a yeast Ty element: evidence for a novel mechanism of gene expression. *Proc. Natl. Acad. Sci. USA* **82**:2829-2833.
- Duyk, G., J. Leis, M. Longiaru, and A. Skalka. 1983. Selective cleavage of the avian retrovirus LTR sequence by the endonuclease associated with the $\alpha\beta$ form of the avian reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **80**:6745-6749.
- Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. deHaseth, A. M. Skalka, and J. Leis. 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. *J. Virol.* **56**:589-599.
- Elbel, H., J. Gafner, A. Stotz, and P. Philippsen. 1981. Characterization of the yeast mobile element Ty1. *Cold Spring Harbor Symp. Quant. Biol.* **45**:609-617.
- Elder, R. T., E. Y. Loh, and R. W. Davis. 1983. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. *Proc. Natl. Acad. Sci. USA* **80**:2432-2436.
- Errede, B., T. S. Cardillo, F. Sherman, E. Dubois, J. Deschamps, and J. M. Wiame. 1980. Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. *Cell* **22**:427-436.
- Errede, B., M. Company, J. D. Ferchak, C. A. Hutchison III, and W. S. Yarnell. 1985. Activation regions in a yeast transposon have homology to mating type control sequences and to mammalian enhancers. *Proc. Natl. Acad. Sci. USA* **82**:5423-5427.
- Farabaugh, P. J., and G. R. Fink. 1980. Insertion of the eukaryotic transposable element Ty1 creates a 5-base pair duplication. *Nature (London)* **286**:352-356.
- Gafner, J., and P. Philippsen. 1980. The yeast transposon Ty1 generates duplications of target DNA on insertion. *Nature (London)* **286**:414-418.
- Garfinkel, D. J., J. D. Boeke, and G. R. Fink. 1985. Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **42**:507-517.
- Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell* **18**:93-100.
- Grandgenett, D. P., and A. C. Vora. 1985. Site-specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. *Nucleic Acids Res.* **13**:6205-6221.
- Kosiba, B. E., B. Errede, T. S. Cardillo, and F. Sherman. 1982. Mutants of yeast overproducing iso-2-cytochrome *c*. *Recent Adv. Yeast Mol. Biol.* **1**:156-172.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning, a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Mellor, J., S. M. Fulton, M. J. Dobson, W. Wilson, S. M. Kingsman, and A. J. Kingsman. 1985. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. *Nature (London)* **313**:243-246.
- Mellor, J., M. H. Malim, K. Gull, M. F. Tuite, S. McCready, T. Dibbayawan, S. M. Kingsman, and A. J. Kingsman. 1985. Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. *Nature (London)* **316**:583-586.
- Panganiban, A. T., and H. M. Temin. 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. *Cell* **36**:673-679.
- Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161-178.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Swanstrom, R., H. E. Varmus, and J. M. Bishop. 1981. The terminal redundancy of the retrovirus genome facilitates chain elongation by reverse transcriptase. *J. Biol. Chem.* **256**:1115-1121.
- Taylor, J. M. 1977. An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes. *Biochim. Biophys. Acta* **473**:57-71.
- Varmus, H. E. 1983. Retroviruses, p. 411-502. *In* J. A. Shapiro (ed.), *Mobile genetic elements*. Academic Press, Inc. New York.